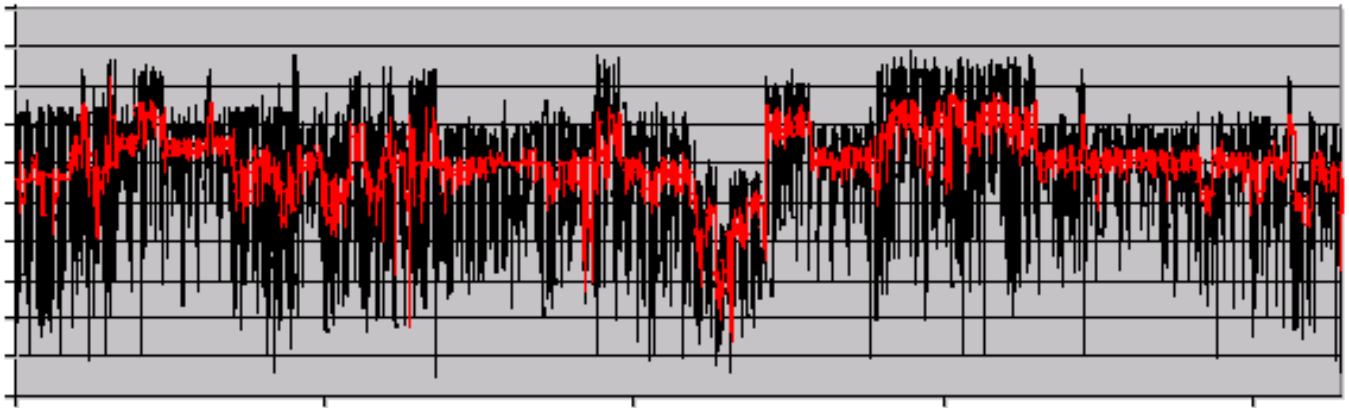


# **Grid Performance Workshop 2004**



**University College London**

**May 12-13, 2004**

**[www.mcs.anl.gov/~jms/Workshop](http://www.mcs.anl.gov/~jms/Workshop)**

# Grid Performance Workshop 2004

<http://www.mcs.anl.gov/~jms/Workshop>

The Grid Performance Workshop 2004 will address the science of performance and the Grid. It will be a forum for identifying and discussing the latest research and current application needs for performance and monitoring data for Grid systems, both in terms of design and use.

Applications are slowly being adapted to run over multiple administrative domains in a coordinated way, but they rarely achieve even a fraction of the possible performance of the underlying systems. In part this is because users very rarely know what performance they could achieve. There is no current data or infrastructure to formulate estimates of baseline performance with which to determine the difference between how an application is currently running and what is possible with a bit of effort. We will begin by focusing on the needs of data-centric scientific applications.

This two-day working meeting will involve a mix of researchers from parallel computing, distributed computing, performance, and Grid areas to discuss the state of the art, what can be leveraged from other work, and what future work is needed. We feel strongly that a mixture of Grid and non-Grid researchers and their background will help advance the understanding of performance and Grids. The result of the workshop will be a white paper on the state of the field and recommendations for future work.

Topics to be addressed include:

- Measurement
  - Baseline performance information
  - Benchmarks - what measurements are needed to represent the needs of the important application classes, and what are the important application classes?
  - Scalability in terms of taking measurements and software infrastructure
  - Measurements, with an emphasis on the statistical appropriateness
- Analysis
  - Gap analysis - how do we determine the gap between the achieved performance, the expected performance, and the possible performance?
  - Scaling of performance metrics
- Simulation - what role does simulation and modeling play in a Grid environment, and how can we verify its appropriateness?
- And the underlying themes of reproducibility, scalability, and usability

We plan to address the performance and the measurements needed, not the actions taken as a consequence of them. This workshop will not discuss the topics of error propagation, fault tolerance, or application algorithm design. We will not discuss architectures or APIs, although at the end of the workshop we hope to know better where standardization might be of use in the future.

# AGENDA

## May 12

9:00-10:00

- Registration and breakfast

10:00-10:45 Introduction and Overview

- Jennifer Schopf, Welcome
- Tony Hey and Stephen Jarvis, A Roadmap for Grid Performance: Reporting From the Frontline

10:45-12:15 Cautionary Tales

- Bill Gropp, How Not to Measure Performance: Lessons from Parallel Computing
- Pete Beckman, Grid Performance, From Linpack to Uptime of a National Grid Infrastructure, a Look at Performance and Metrics for Production Grid

Lunch

1:30-3:00 Application Classes and the Measurements Needed for Them

- Ron Perrot, GT2 and GT3 Performance Timings for Projects at the Belfast eScience Centre
- Ani Thakur, Tracking Database Usage and Measuring Data Mining Performance in the SDSS
- Joel Saltz, Cooperative Biomedical Research, Data Virtualization and Grid Computing

3:30-5:00 Breakout - Measurements

- What do we have? What do we need?

5:00-5:30

- Summary of breakouts

6:30

- Banquet at Navarro's, 67 Charlotte Street, London W1T 4PH

## May 13

8:00-9:00

- Breakfast

9:00-10:30 Performance Evaluation and Benchmarking Today

- Brian Tierney, Grid Troubleshooting Using the Netlogger
- Jack Dongarra, A Look at Some Ideas and Experiments
- Allan Snavely, Benchmark Probes for Grid Assessment

11:00-12:30 Science of field measurements applied to the Grid

- Les Cotrell, Network Monitoring Today: Why, How, Challenges, Infrastructures, Federations and the Grid
- Graziano Obertelli, Observing Resources in the Wild
- followed by a 30 min discussion of the topic

Lunch

1:30-3:30 - Breakout

- What is the specific data we need to collect?

3:30-4:00

- Workshop Summary (with snacks)

# SPEAKERS AND ABSTRACTS

## ***Introduction and Overview***

**Jennifer M. Schopf, Argonne National Laboratory**  
**Welcome**

**Tony Hey, Director e-Science Programme, EPSRC**  
**Stephen Jarvis, University of Warwick, Midlands e-Science Centre**  
**A Roadmap for Grid Performance: Reporting From the Frontline**

After a brief review of the context of the UK e-Science Program and of JISC e-Infrastructure funding, the relevance of performance issues for a number of UK projects will be reviewed. Rather than focus on the intricacies of any e-Science project in particular, this talk will introduce open performance issues, with a focus on identifying the performance properties and requirements of the application, of the target architectures and of the supporting operating system (Grid middleware) components.

Suggested reading:

- “Performance-based Middleware for Grid Computing”, G.R. Nudd, S.A. Jarvis, to appear in *Concurrency and Computation: Practice and Experience*, 2004.  
[www.dcs.warwick.ac.uk/~saj/papers/CPE.pdf](http://www.dcs.warwick.ac.uk/~saj/papers/CPE.pdf)
- “The Impact of Predictive Inaccuracies on Execution Scheduling”, S.A. Jarvis, L. He, D.P. Spooner, G.R. Nudd, to appear *Performance Evaluation: An International Journal, special issue on Performance Modeling and Evaluation of High-performance Parallel and Distributed Systems*, 2004.  
[www.dcs.warwick.ac.uk/~saj/papers/PEpaper.pdf](http://www.dcs.warwick.ac.uk/~saj/papers/PEpaper.pdf)
- “ARMS: An Agent-Based Resource Management System for Grid Computing”, J. Cao, S.A. Jarvis, D.J. Kerbyson, G.R. Nudd, *Scientific Programming (Special issue on Grid Computing)*, 10(2):135-148, 2002. [www.dcs.warwick.ac.uk/~saj/papers/arms.pdf](http://www.dcs.warwick.ac.uk/~saj/papers/arms.pdf)

## ***Session 1: Cautionary Tales***

**William Gropp, Argonne National Laboratory**  
**How Not to Measure Performance: Lessons from Parallel Computing**

Performance measurement and modeling has a long history in parallel computing. While it is relatively easy to perform measurements, it is much harder to perform measurements that are scientifically valid and provide the basis for performance prediction. While grids and parallel computers have important differences, there are enough similarities that some of the lessons from parallel computing can be applied to grid performance, particularly with respect to communications, remote I/O, and end-to-end applications performance (e.g., rather than raw network performance). This talk will review some of

the successes and failures in parallel performance measurement. Differences between grids and parallel computers will also be used to suggest approaches suitable for grid measurements.

Suggested reading:

- Performance Engineering Workshop, Dec 9-10, 2002, e-Science Institute, Edinburgh
- <http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=127>
- “Reproducible Measurements of MPI Performance Characteristics”, William Gropp and Ewing Lusk, in the Proceedings of *PVMMPI'99*  
<http://www.mcs.anl.gov/~gropp/bib/papers/1999/pvmmmpi99/mpptest.pdf>
- “Twelve Ways to Fool the Masses When Giving Performance Results on Parallel Computers”, David H. Bailey, *Supercomputing Review*, Aug. 1991,  
<http://crd.lbl.gov/~dhbailey/dhbpapers/twelve-ways.pdf>

### **Pete Beckman, TeraGrid Project Grid Performance, From Linpack to Uptime of a National Grid Infrastructure, a Look at Performance and Metrics for Production Grid**

Supercomputing over the years has developed a number of metrics for system performance, from the raw measures of memory bandwidth and message passing to application benchmarks to system availability and uptime. Can Grid systems evolve meaningful benchmarks that users can easily understand and correlate with their applications and experiences? Moreover, can they be used for "Service Level Agreements" (SLAs)? As Grid systems evolve toward production infrastructures, can measures of "up" and "minimum performance" eventually be written into funding contracts similar to what is specified for supercomputer procurements? In this discussion we present some of the metrics and benchmarks that the TeraGrid is exploring for operating its Production Grid, and the success of deploying them.

### ***Session 2: Application Classes and the Measurements Needed for Them***

#### **Ron H. Perrot, Belfast eScience Centre, The Queen's University of Belfast GT2 and GT3 performance timings for Projects at the Belfast eScience Centre**

The presentation will cover three applications. First, we will discuss an academic application in the field of computational physics that has been running as a GT2 application since March 2003 on the UK L2G. The application's long-term performance over some 12 months will be presented. Belfast eScience centre has focused on the industrial applications of grid and web service technology and two industrial projects are

presented. The GridCast project is developing grid services for a media broadcast infrastructure. This application will be used to illustrate scenario based timings for GT3 services running, since September 2003, in a wide area GT3 testbed. The GEDDM project is prototyping services for high-performance data mining. This will be used to illustrate the performance of services within a high-throughput commercial data mining engine.

Suggested reading:

- Rmatrix, <http://www.qub.ac.uk/escience/projects/rmatrix>
- GridCast Project, <http://www.qub.ac.uk/escience/projects/gridcast>
- GEDDM (Grid Enabled Distributed Data Mining),  
<http://www.qub.ac.uk/escience/projects/geddm>

## **Joel Saltz, The Ohio State University**

### **Cooperative Biomedical Research, Data Virtualization and Grid Computing**

Collaborative biomedical research increasingly requires coordinated generation and analysis of large datasets generated through the use of many types of imaging modalities. Both logistical and medico-legal requirements dictate that image data be maintained in a distributed manner. Development of new techniques for standardized image interpretation and image feature detection requires that grid middleware be able to invoke algorithms on user defined sets of images. In our view, this application area also poses specific requirements on how grid systems should manage metadata.

In my talk, I will describe issues associated with support for emerging Grid based cooperative biomedical research enterprises and also outline how some of these requirements can be met by the DataCutter, STORM and Mobius grid middleware systems.

Suggested reading:

- “Image Processing on the Grid: A Toolkit for Building Grid-enabled Image Processing Applications”, Shannon Hastings, Tahsin M. Kurç, Stephen Langella, Ümit V. Çatalyürek, Tony Pan, Joel H. Saltz, *Proceedings of CCGRID 2003*, 2003.  
<http://medicine.osu.edu/informatics/publicationsPages/pdf/TR16.pdf>
- “Database Support for Data-Driven Scientific Applications in the Grid”, Sivaramakrishnan Narayanan, Tahsin M. Kurç, Ümit V. Çatalyürek, Joel H. Saltz. *Parallel Processing Letters* 13(2): 245-271 (2003)  
<http://medicine.osu.edu/informatics/publicationsPages/pdf/TR10.pdf>
- “Active Proxy-G: Optimizing the Query Execution Process in the Grid”, Henrique Andrade, Tahsin M. Kurç, Alan Sussman, Joel H. Saltz. *Proceedings of SuperComputing 2002*, 2002.  
<http://medicine.osu.edu/informatics/publicationsPages/pdf/sc02-tr.pdf>

## **Ani Thakur, The Johns Hopkins University**

### **Tracking Database Usage and Measuring Data Mining Performance in the SDSS**

The Sloan Digital Sky Survey is now serving 2 TB of data online via the SkyServer websites and the SkyQuery federated query Web Service. Measurement of the data throughput and query performance is an important part of the archive operations and site maintenance. I will discuss the extensive logging and traffic analysis that we are currently doing on the SkyServer sites. I will also describe our ongoing efforts to provide scalable high-speed access to SDSS data through spatial indexing, separating the batch and interactive servers, data striping, vertical partitioning and horizontal partitioning across a cluster of data nodes. We are using a cluster-finding algorithm as a test problem to compare performance between the Terabyte Analysis Machine (a Grid prototype) and partitioned database access on our SDSS cluster. We plan to incorporate the data partitioning into an automated workflow data import pipeline that we have developed at JHU. Finally, I will discuss our effort to devise a logging strategy and data model for the Virtual Observatory.

Suggested reading:

- “Migrating MultiTerabyte Archive from Object to Relational Databases”, A. R. Thakar, A.S. Szalay, P. Z. Kunszt, and J. Gray, *Computing in Science and Engineering*, 5, 5 (Sep/Oct 2003). <http://www.sdss.jhu.edu/sx/pubs/cise03.pdf>
- “The Terabyte Analysis Machine Project, The Distance Machine: Performance Report”, J. Annis, G. Garzoglio, C. Stoughton, *Computing in High Energy and Nuclear Physics (CHEP)*, 2001. [http://projects.fnal.gov/act/tam/ANN\\_report.htm](http://projects.fnal.gov/act/tam/ANN_report.htm)
- “High-Speed Access for an NVO Data Grid Node”, M. Nieto-Santisteban, A. R. Thakar, A.S. Szalay, and T. Malik, *T. AISRP 2003 PI Workshop Proceedings*, 2003. [http://trotsky.arc.nasa.gov/~pgazis/News/workshop2003/Nieto\\_Santisteban/AISRP2003.ppt](http://trotsky.arc.nasa.gov/~pgazis/News/workshop2003/Nieto_Santisteban/AISRP2003.ppt)
- “Data Mining the SDSS SkyServer Database”, J. Gray, D. Slutz, A. Szalay, A. Thakar, J. vandenBerg, P. Kunszt, and C. Stoughton, *Microsoft Research Technical Publication*, MSR-TR-2002-01, 2002. [http://research.microsoft.com/research/pubs/view.aspx?msr\\_tr\\_id=MSR-TR-2002-01](http://research.microsoft.com/research/pubs/view.aspx?msr_tr_id=MSR-TR-2002-01)

## **Session 4: Performance Evaluation and Benchmarking Today**

### **Brian L. Tierney, Lawrence Berkeley National Lab**

#### **Grid Troubleshooting Using the NetLogger**

Monitoring is the measurement and publication of the state of a computing/networking component at a particular point in time. To be effective, monitoring must be "end-to-end", meaning that all components between the application endpoints must be monitored. This includes software (e.g., applications, services, middleware, operating systems), end-host hardware (e.g., CPUs, disks, memory, network interface), and networks (e.g., routers, switches, or end-to-end paths).

Monitoring is required for a number of purposes, including status checking, troubleshooting, performance tuning, debugging, application steering, characterizing usage, planning, setting expectations, developing and auditing service level agreements. For example, assume a Grid job has been submitted to a resource broker, which uses a reliable file transfer service to copy several files to the site where the job will run, and then runs the job. This particular process should normally take 15 minutes to complete, but two hours have passed and the job has not yet completed. Determining what, if anything, is wrong is difficult and requires a great deal of monitoring data. Is the job still running or did one of the software components crash? Is the network particularly congested? Is the CPU particularly loaded? Is there a disk problem? Was a software library containing a bug installed somewhere? Monitoring provides the information to help track down the current status of the job and locate problems.

This talk will focus on how the NetLogger Toolkit (<http://dsd.lbl.gov/NetLogger/>) can be used to address the problem of Grid Troubleshooting. NetLogger is an open source collection of tools that include instrumentation libraries, monitoring data visualization tools, and monitoring data collection and archival tools. The talk will also cover the current "hard problems" in Grid troubleshooting.

Suggested reading:

- Netlogger, <http://dsd.lbl.gov/NetLogger/>
- “On-Demand Grid Application Tuning and Debugging with the NetLogger Activation Service”, Dan Gunter, Brian L. Tierney, Craig E. Tull, and Vibha Virmani, *Proceedings of the 4th International Workshop on Grid Computing (Grid2003)*, LBNL-52991, 2003. <http://www.didc.lbl.gov/papers/ActivationService.Grid2003.pdf>
- “Enabling Network Measurement Portability Through a Hierarchy of Characteristics”, Bruce B. Lowekamp, Brian Tierney, Les Cottrell, Richard Hughes-Jones, Thilo Kielmann, and Martin Swany, *Proceedings of the 4th International Workshop on Grid Computing (Grid2003)*, LBNL-53013, 2003. <http://www.didc.lbl.gov/papers/Grid2003-characteristics-hierarchy.pdf>
- “Dynamic Monitoring of High-Performance Distributed Applications”, Dan Gunter, Brian Tierney, Keith Jackson, Jason Lee, and Martin Stoufer, *Proceedings of the 11th IEEE Symposium on High Performance Distributed Computing (HPDC-11)*, 2002, LBNL-49698 <http://www.didc.lbl.gov/papers/HPDC02-HP-monitoring.pdf>

## **Jack Dongarra, University of Tennessee and Oak Ridge National Lab**

### **A Look at Some Ideas and Experiments**

In this talk we will look at some of the general ideas of performance evaluation in the Grid and some experiments that have been conducted using Grid resources.

Suggested reading:

- “SRS - A Framework for Developing Malleable and Migratable Parallel Applications for Distributed Systems”, Sathish Vadhiyar and Jack Dongarra, *in Review for Parallel Processing Letters*, 2003.



<http://www.netlib.org/utk/people/JackDongarra/PAPERS/vadhiyar-framework-malleable.pdf>

- “Scheduling in the Grid Application Development Software Project”, Holly Dail, Otto Sievert, Fran Berman, Henri Casanova, Asim YarKhan, Sathish Vadhiyar, Jack Dongarra, Chuang Liu, Lingyun Yang, Dave Angulo, and Ian Foster, in “Grid Resource Management”, ed. Nabrzyski, Schopf and Weglarz, Kluwer, 2003.  
<http://www.netlib.org/utk/people/JackDongarra/PAPERS/grads-kluwer2003.pdf>
- “NetSolve: Past, Present, and Future - a Look at a Grid Enabled Server”, Sudesh Agrawal, Jack Dongarra, Keith Seymour, and Sathish Vadhiyar, in *Grid Computing: Making the Global Infrastructure a Reality*, eds F. Berman, G. Fox, and A. Hey, 2003.  
<http://www.netlib.org/utk/people/JackDongarra/PAPERS/netsolve-ppf.pdf>

**Allan Snaveley, UCSD/SDSC**

### **Benchmark Probes for Grid Assessment**

Like all computing platforms, grids are in need of a suite of benchmarks by which they can be evaluated, compared and characterized. As a first step towards this goal, we have developed a set of probes that exercise basic grid operations with the goal of measuring the performance and the performance variability of basic grid operations, as well as the failure rates of these operations. We present measurement data obtained by running our probes on a grid testbed that spans 5 clusters in 3 institutions. These measurements quantify compute times, network transfer times, and Globus middleware overhead. Our results help provide insight into the stability, robustness, and performance of our testbed, and lead us to make some recommendations for future grid performance

Suggested reading:

- “Benchmark Probes for Grid Assessment”, Greg Chun, Holly Daily, Henri Casanova and Allan Snaveley, *Proceedings of the IPDPS workshop on Grid Performance*, 2004.  
[http://grail.sdsc.edu/papers/grasp\\_hpgc\\_04.pdf](http://grail.sdsc.edu/papers/grasp_hpgc_04.pdf)
- “Benchmarks for Grid Computing: A Review of Ongoing Efforts and Future Directions”, Allan Snaveley, Greg Chun, Henri Casanova, Rob F. Van der Wijngaart, and Michael A. Frumkin, *SIGMETRICS Performance Evaluation Review (PER)*, Vol. 30, No. 4, March 2003. [www.sdsc.edu/pmac/Papers/sigmetrics\\_final.pdf](http://www.sdsc.edu/pmac/Papers/sigmetrics_final.pdf)

## ***Session 5: Science of Field Measurements Applied to the Grid***

**R. Les Cottrell, SLAC, Stanford University**

### **Network Monitoring Today: Why, How, Challenges, Infrastructures, Federations and the Grid**

This talk will first cover why we monitor networks (with some examples), and how we monitor them using active and passive methods. We will also describe some of the challenges network monitoring faces today. We will then discuss and compare some of the Network Monitoring Infrastructures (NMIs) in place today and the challenges they face. Following this I will briefly describe a proposal (MAGGIE) to create a Federation

of NMIs together with its goals. Finally we will discuss how the measurements may be useful to the grid and how they may be incorporated into Grid applications.

Suggested reading:

- "Internet Research Needs" presented to LSN, [www.slac.stanford.edu/grp/scs/net/talk03/lsn-jun03.ppt](http://www.slac.stanford.edu/grp/scs/net/talk03/lsn-jun03.ppt)
- "GGF Network Monitoring Working Group", <http://www-didc.lbl.gov/NMWG/>
- "The PingER Project: Active Internet Performance Monitoring for the HENP Community", W. Matthews, R. L. Cottrell, *IEEE Communications Magazine*, May 2000, <http://www-iepm.slac.stanford.edu/paperwork/ieee/ieee.pdf>
- "ABwE: A Practical Approach to Available Bandwidth Estimation", J. Navratil, R. L. Cottrell, *Proceedings of Passive and Active Measurements*, 2003, <http://moat.nlanr.net/PAM2003/PAM2003papers/3781.pdf>
- "Experiences and Results from a New High Performance Network and Application Monitoring Toolkit", R. L. Cottrell, C. Logg, I-Heng Mei, SLAC-PUB-9641, *proceedings of PAM 2003*, April 2003, <http://moat.nlanr.net/PAM2003/PAM2003papers/3768.pdf>

## **Graziano Obertelli, University of California Santa Barbara Observing Resources in the Wild**

The Grid promises the ability to harness more computational and storage capabilities than has ever before been possible, but there is a catch. Each resource must be evaluated in terms of the performance it can deliver, and when it can deliver that performance, or applications will simply be lost in a Maelstrom of dynamically changing heterogeneity. That is, the power available from Grid resources fluctuates, often wildly, and either the Grid infrastructure or the applications themselves must be able to mitigate and/or tolerate this fluctuation for the true computing potential of the paradigm to be realized.

To bring order out of the chaos that the dynamic federation and sharing of resources engenders, application schedulers and users must be able to determine the "value" of each resource to the application or user that will access it. To do so requires measurements of the application-deliverable performance. However, gathering and managing performance measurements scalably and reliably from a fluctuating, potentially global pool of Grid resources poses some unique scientific and engineering challenges.

In this talk we will present many of the difficulties associated with grid performance measurement and potential solutions we have explored as part our work with the Network Weather Service (NWS). The NWS is a scalable and robust system for gathering, managing and forecasting grid resource performance and availability. We will focus on the performance demands that a system for serving dynamically changing performance data must meet, and we will discuss our experiences with measuring several different grid resource characteristics in a variety of settings.

# Thanks

We gratefully thank NSF Division of Shared CyberInfrastructure and the Division of Computer and Network Systems, Microsoft, and the Joint Information Systems Committee (JSIC) for their help in supporting this meeting. We are also tremendously grateful to Pete Clarke and the UCL crew for their assistance in local arrangements and planning.

## **Workshop Steering Committee:**

- Jennifer M. Schopf, Argonne National Laboratory
- Mark Baker, University of Portsmouth
- Fran Berman, UCSD
- Jack Dongarra, University of Tennessee
- Ian Foster, Argonne National Laboratory and the University of Chicago
- Bill Gropp, Argonne National Laboratory
- Tony Hey, UK e-Science Centre

Contact for details: Jennifer M. Schopf [jms@mcs.anl.gov](mailto:jms@mcs.anl.gov)



Map 3 General area around UCL

UCL Massey Theater

Tavistock Hotel

